Data Analysis | Programming Challenge

Overview:

This programming challenge consists of two tasks. First task requires you to develop a backend engine that will expose a web API. These APIs can be used by client to search the data based on a specific criteria. In the second part, you will be required to do data clustering to organize data in meaningful groups.

Please note that you are required to use Python version 3.x.x throughout this task. Choice of other supporting frameworks and libraries is up to you.

Submission:

You are required to submit this task using github. Please create a profile on github If you don't already have one. Upload all the code, graphs and results to a github repository and share the link to repository in the email.

You will have **one week** to submit this assignment.

Scoring:

Primary scoring will be done on the correctness of your code and output. However, We will be considering several other factors for evaluation e.g.

- Code readability
- Performance
- Documentation
- Error handling
- Test cases
- Evaluation/accuracy scores.

You are encouraged to improvise on the above list, as this is only for the hint.

Task-01:

This task requires you to crawl data from French Chamber of Commerce. You should Collect at-least 50 data points(company profiles). The final output should be a json File named *companies.json* which should contain data in the following format

```
{
  "uid": "http://www.fccsingapore.com/membership/directory/account/1271",
  "company_name": "Invest Expat",
  "company_description": "Neque porro quisquam est qui dolorem ipsum quia dolor",
  "company_industry": "wealth management",
  "company_address": "21/F On Hing Building1 On Hing TerraceCentral HONG KONG",
  "company_website": "http://www.invest-expat.com"
}
```

Please upload your crawling code along with the data file to github.

Task-02:

In this task you are required to do data analysis. Please find *data.csv* file provided. It contains data for 20,000 companies. Column description is as follows:

Id: Unique identifier for company.
Name: Company Name.
Country: Country where the company is located.
Employees: Number of Employees.
Ownership: Ownership status of company.
Revenue: Revenue of company.
Industry: Industry Classification of company.

You are required to analyze the data to find correlations and perform EDA on it. Present all your analysis and graphs in a jupyter notebook.